Summary

Introduction

Dealing with heterogeneity among study treatment effects, or "the situation in which differences in study outcomes are not readily accounted for by sampling variation," (Colditz GA, 1995) is one of the most important challenges facing a meta-analyst. The National Center for Complementary and Alternative Medicine (NCCAM) recognized the importance of heterogeneity in meta-analysis. With the Agency for Healthcare Research and Quality (AHRQ), NCCAM established the objective of this study to compare and contrast several strategies for understanding heterogeneity via meta-regression methods. They asked the Southern California Evidence-Based Practice Center, in its role as technical support to NCCAM, to conduct the study and to produce this report.

If heterogeneity is found or suspected to exist, the common approaches used in meta-analysis are to:

- Stratify the studies into homogeneous subgroups and then fit a separate fixed effects estimate, e.g., of the pooled odds ratio, in each strata.
- Construct a random effects estimate across all studies. A random effects approach incorporates both within-study and between-study variability. We note that some argue that if heterogeneity exists among studies, a summary measure across those studies should not be provided.
- Fit a meta-regression model that explains the heterogeneity in terms of study-level covariates.

The broad objective of this report is to focus on this strategy of modeling the differences between studies, by comparing and contrasting several meta-regression methods.

Methodology

We conducted a systematic review of MEDLINE®, HealthSTAR, EMBASE, MANTIS, SciSearch®, Social SciSearch®, Allied and Complementary Medicine, the Current Index to Statistics, and the Methodology Register of the Cochrane Library through March 2001 using the search terms "metaregress-" or "meta" within two words of "regress-" in order to identify publications on meta-regression. We supplemented these searches with articles identified by experts, and by searching the reference lists of relevant articles.

Given the variety of meta-regression approaches available, our first analytic objective was to propose a common statistical framework, using the knowledge gained from the articles found via our systematic review, in which all meta-regression models could be expressed.

We implemented a simulation study to compare the different meta-regression modeling approaches. Simulation allowed us to set up a scenario (the "true" model), simulate data from that model, estimate parameters using various meta-regression models, and then compare the estimated parameters of each model in terms of bias.

We convened a one-day meeting of nine experts on heterogeneity in meta-analysis, and meta-regression. Prior to the meeting, the experts were sent background materials, including the preliminary parameters we were considering for our simulation study. The experts were also asked to suggest additional meta-regression publications. During the meeting, four of the experts presented half-hour talks spanning different types of meta-regression approaches.

The experts reached agreement on the parameters needed to complete the simulation, and additional analyses to conduct. The meeting was audio-taped and transcribed to assist in the preparation of this report.

Findings

The systematic review produced 85 publications relevant to meta-regression. We categorized the publications into seven categories based on the primary focus of the article. The first four categories were the main meta-regression methods: fixed effects models (4 publications); random effects models (11 publications); control rate models (9 publications); and Bayesian and/or hierarchical models (13 publications). We also defined an "overview" category that contained articles that surveyed meta-regression methods and/or focused on the unique challenges of such a modeling effort, including for example discussion of ecological bias (19 publications). Our sixth category consisted of articles that addressed modeling studies that had multiple treatment arms and/or multiple endpoints or outcomes, as such studies present unique challenges to the meta-analyst (5 publications). Our seventh category consisted of examples (24 publications).

Using the knowledge gained from our review of the retrieved articles, we proposed a common statistical framework in which all meta-regression methods could be expressed. We restricted attention to dichotomous outcomes only. We discussed scenarios in which meta-regression might be informative, and presented the common meta-regression approaches using our proposed notation.

Our expert panel made several recommendations regarding the simulation parameters. The panel also generally identified the need for outreach by the methodological community to the user community in advising how to conduct, interpret, and present meta-regression analyses, including the development of software and diagnostic aids to assess models.

In our simulation study, we evaluated five meta-regression methods: fixed effects with and without covariates; random effects with and without covariates; and control rate meta-regression. The simulation was a complete factorial design including all possible 7,776 combinations of the simulation parameters.

We compared methods in terms of bias in the estimation of the additive treatment effect, which is the parameter typically estimated in meta-analyses. The results were evaluated using an analysis-of-variance (ANOVA) model relating the simulation parameters to the bias. Across the five different meta-regression methods, six terms in a three-way ANOVA model were found to be practically important as they captured contributions to the bias of 10% or greater on average.

Conclusions from a Statistical Perspective

Meta-regression methods will be increasingly used. Their attractiveness lies in their potential to explain differences between studies, thereby helping the clinician and decision-maker determine when, where, and for whom a treatment is beneficial. Our expert panel noted the usefulness and timeliness of this report.

Our panel had several general recommendations regarding meta-analysis and meta-regression. Foremost, the panel echoed the guidance given by others: measuring and incorporating heterogeneity in a meta-analysis is not sufficient. Meta-analysts should investigate and attempt to understand the causes of heterogeneity. The panel identified the need for outreach by the methodological community to the user community in advising how to conduct, interpret, and present meta-regression analyses, including the development of software and diagnostic aids to assess models.

The panel made several recommendations that we were able to include in this report. Some recommendations are delegated to future research. The panel also addressed the next methodological topic for the Southern California Evidence-Based Practice Center given our role as technical support to NCCAM. The panel recommended that if we undertake as our next methodological topic the quality assessment of observational studies, we focus on a specific clinical topic as a "case study." The panel recommended against developing a global scale, and also did not advise considering observational study quality in general.

Our simulation results produced the following guidelines for the meta-regression practitioner:

- In general, failure to incorporate important covariates at either the study or person level, can bias the results of a meta-analysis.
- Despite the importance of including covariates, a model that includes a covariate that is an aggregate of a person-level characteristic rather than a study characteristic can produce biased results. The trade-off between the biases of incorporating an aggregated covariate versus excluding it requires further exploration.
- If the control rate affects treatment, the meta-analysis should incorporate the control rate. However, control rate meta-regression is susceptible to bias via the correlation between the control rate and other omitted covariates. This suggests that extensions of control rate meta-regression to include other covariates may prove useful.
- As always, larger number of studies and larger number of patients per study can reduce bias with proper modeling.

In summary, our key message to practitioners is they should explore the causes of heterogeneity via the inclusion of covariates at both the person level and study level. Either fixed effects or random effects methods can be used to support this exploration. Note that our work presented in this report has not addressed confidence interval construction and statistical significance testing. Further work in this dimension may reveal differences between fixed and random effects approaches.

Conclusions from a Nonstatistical Perspective

Consider a meta-analysis of randomized controlled trials of a treatment to reduce heart disease mortality. Assume that the study-level variable aggregated from person-level data is the average disease severity, e.g., average blood pressure, among persons in each specific trial. Assume further that another study-level variable, whether the trial occurs in a hospital versus an outpatient setting, is also available. The control rate in this example is the rate of heart disease mortality in the control group in each trial.

Based on our simulation study, our first conclusion stated above is *"In general, failure to incorporate important covariates at either the study or person level, can bias the results of a meta-analysis."* Consider in our example the possibility that the treatment effect may not be the same for mild and severe cases. The treatment may have no effect for patients with very high blood pressure (high severity), while it has a strong effect for patients with mildly elevated blood pressure (low severity). Further we would anticipate that a hospital-based trial may in general accumulate sicker patients than an outpatient-based trial. Failure to account for the trial to trial variation in patient severity may lead to the incorrect conclusion that the treatment is less effective in hospital settings than in outpatient settings.

Our second conclusion is *"Despite the importance of including covariates, a model that includes a covariate that is an aggregate of a person-level characteristic rather than a study characteristic can produce biased results. The trade-off between the biases of incorporating an aggregated covariate versus excluding it requires further exploration."* In many statistical problems, it has been observed that using average quantities in place of person-specific quantities can lead to biased results and erroneous conclusions. Many of these problems fall under the general label of "ecological fallacy." Although further research is required, it may be the case that using aggregated variables in meta-analysis, such as average blood pressure, is a useful first step in understanding how the treatment effect differs across different types of patients and settings. However, we must always bear in mind that such conclusions should be considered exploratory rather than confirmatory. When interesting findings are discovered in this manner, person-level data from large trials may be required for confirmation.

Our third conclusion is "*If the control rate affects treatment, the meta-analysis should incorporate the control rate. However, control rate meta-regression is susceptible to bias via the correlation between the control rate and other omitted covariates. This suggests that extensions of control rate meta-regression to include other covariates may prove useful.*" We think that the control rate meta-regression method holds great promise to help us understand the relationship between treatment effect, illness severity and differences in trial protocols. However, the success of control rate meta-regression by its very nature begs the question of what other covariates might help us explore this heterogeneity. With time, we believe improved methods will be available to address these questions. In the meantime, where control-rate meta-regression differs from a simpler meta-regression, policy conclusions should be made tentatively and with caution.

Our fourth conclusion is *"As always, a larger number of studies and larger number of patients per study can reduce bias with proper modeling."* The user of meta-regression should remember that the degrees of freedom he or she has to understand study characteristics are

severely limited by the number of studies. Larger individual studies will always be a scientific gold standard that cannot be completely replaced by the meta-analysis of smaller studies.

Alternative Medicine Meta-analysis

NCCAM and AHRQ were especially interested in the topic of heterogeneity in meta-analysis given its relevance to alternative medicine literature. The challenges faced in synthesizing this literature are very different from those faced in for example the cardiovascular literature. The latter consists mainly of large randomized controlled trials. In contrast, the alternative medicine literature consists mainly of small trials, and nonrandomized studies. In addition to study design heterogeneity, the interventions are heterogeneous as well as the patient populations. Thus, methods for dealing with heterogeneity are particularly relevant.

Staff of the Southern California Evidence-Based Practice Center have applied meta-regression in the alternative medicine setting. In a systematic review of the evidence on ephedra and ephedrine, we used meta-regression to compare weight loss efficacy between groups receiving ephedrine; ephedrine plus caffeine; and ephedra plus herbs containing caffeine versus placebo. In a meta-analysis of spinal manipulation, we developed meta-regression models for acute and chronic back pain patients predicting short-term and long-term pain and function. These models took the unique approach of denoting the spinal manipulation group as the comparison group against which all other treatments, such as sham or physical therapy, were compared. The usual strategy would be to compare versus placebo or control. The knowledge gained via the research presented in this report impacted the application of meta-regression to these alternative medicine questions, and improved our ability to synthesize and understand these therapies.